

# COMPARING THE APPLICABILITY OF BOX-JENKINS ARIMA METHODOLOGY AND ARCH/GARCH METHODOLOGY AMONG REAL DATA SETS

M.G.S.M.A. Ferdinandis

(8101)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree  
Master of Science

Department of Mathematics

University of Moratuwa

Sri Lanka

February 2011

University of Moratuwa



96816

96816



96816

51 "11"

51 (043)

## DECLARATION

"I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any University or other institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text"

Shanika Ferdinandis

Signature:

Date: 08/03/2011


"I hereby grant the University of Moratuwa the right to archive and to make available my thesis or dissertation in whole or part in the University Libraries in all forms of media, subject to the provisions of the current copyright act of Sri Lanka. I retrain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation."

Shanika Ferdinandis

Signature:

Date: 08/03/2011

"I have supervised and accepted this thesis/dissertation for the award of the degree"

Signature of the supervisor: 

Date 02/02/2011

## ACKNOWLEDGMENTS

I would like to express my special thanks and gratitude to my supervisor, Mr.T.M.J.A Cooray, Senior Lecturer Department of Mathematics, University of Moratuwa for his expertise and assistance in clarifying questions regarding time series analysis and the advise he gave which was invaluable in enabling me to compute this thesis.

A special appreciation to the Head of the Department of Mathematics Dr. M.Z.M Malhardeen , Course Co-ordinator Dr. T.S.G Peiris, all the staff members of the Mathematics Department, University of Moratuwa and my colleagues at the university who provided me with advice and support throughout the writing of the thesis.

Last, but not least I would like to thank my ammi and thatththi for their ongoing support and encouragement throughout my studies, particularly in the completion of this thesis.

## Abstract

The main objective of this research is to compare the applicability of Box-Jenkins ARIMA methodology and ARCH/GARCH methodology among two real data sets. This study addresses the question of how to analyze time series data, identify structures, explain behaviours, model the identified structures and using the insight gained, to analyze and forecast values for the specific time series. For the purpose of this study the time series data included, the total kurakkan yield obtained from the Census department of Sri Lanka and the Money series obtained from the International Financial Statistics data source of Central bank of Sri Lanka. Each of the time series has its own characteristics and different methodologies were needed to require a deeper understanding of the time series data.

The analysis of time series constitutes an important area of statistics. The kurakkan yield data set consisted of a few missing values. Three different approaches namely deterministic, stochastic and state space method were used to estimate these missing values. Out of the three approaches the state space method gave the best estimates. Once the missing values were fitted the complete series was used to analyze and then forecast values. To build the models and perform the analysis a statistical software called "Minitab" and the software package called "E-views" was used. The best model obtained was a seasonal ARIMA model with 2 non-seasonal AR terms and 2 non-seasonal MA terms with one seasonal differencing. The model was used to forecast values and the accuracy measure MAPE was 1.65% for the ARIMA model fitted, which was the minimum value of MAPE for all the en-bloc methods mentioned above. The errors of this model were independent and identically distributed and followed a normal distribution.

The main difference between the two time series data sets used for this study is that the money series obtained is a high-volatile data series which includes heteroscedasticity. For this data series the ARIMA methodology cannot be used since the data will not become stationary to fit a model. Therefore the ARCH/GARCH methodology was used to deal with the money data series. To build models for this series the software package called "E-views" has been used. Different ARCH models and GARCH models were fitted to this data set and the parameters were chosen so that the kurtosis value was closer to three. The best model was, a logarithmic transformation of the money series with one GARCH term and no ARCH terms. This model yielded a kurtosis value of 3.09. The main model for this data set did not include any AR or MA terms. However a very large number of data points are required to model the series with AR and MA terms in the main model.

## TABLE OF CONTENTS

Declaration of the candidate & Supervisor	i
Acknowledgements	ii
Abstract	iii
Table of content	iv
List of Figures	x
List of Tables	xiv
List of abbreviations	xv
Chapter 1: Introduction	
1.1 Nature of Time Series	1
1.2 Analysis of time series using En-bloc methods	3
1.3 Box-Jenkins approach for ARIMA methodology	5
1.4 Nature of high-volatile Time Series	7
1.5 Missing values in Time Series data	8
1.6 Aim of research	10
1.7 Significance of research	10
1.8 Data used for the study	11
1.9 Outline of research	11
Chapter 2: The traditional Quantitative forecasting methods	
2.1 Introduction	13

2.2	Decomposition	13
2.2.1	Additive Decomposition	14
2.2.2	Multiplicative Decomposition	14
2.2.3	Advantages and Disadvantages of the Decomposition method	15
2.3	Exponential Smoothing Method	16
2.3.1	Single Exponential Smoothing	16
2.3.2	Holt's Linear Method	17
2.3.3	Holt-Winters' Method	17
2.3.4	Advantages and Disadvantages of Exponential Smoothing Methods	18
2.4	Box-Jenkins ARIMA Methodology	19
2.4.1	Examining correlation in Time Series data	20
2.4.2	The White noise model	20
2.4.3	Sampling distributions of Auto Correlations	20
2.4.4	Box-Pierce Q-Statistic	21
2.4.5	The Partial Auto Correlation	21
2.4.6	Recognizing seasonality	22
2.4.7	Examining Stationarity	23

2.4.8	Differencing	23
2.4.9	Variance stabilizing transformations	24
2.4.10	Seasonal Differencing	25
2.4.11	ARIMA ( p,d,q) Model	25
2.4.12	Seasonality and ARIMA Models	28
2.4.13	Akaikes' Information Criterion	29
2.5	Measures of forecast performances	29

Chapter 3: Methods for the estimation in missing values in time series

3.1	Introduction	32
3.2	Deterministic Modelling	32
3.2.1	Least Squares Approximation	33
3.2.2	Interpolating with a Cubic Spline	34
3.3	Stochastic Modelling	35
3.3.1	Box-Jenkins Modelling and missing values	36
3.4	State Space Modelling	37
3.4.1	Analysis of missing data and the Expectation Maximization (EM) Algorithm	32
3.5	Data used for the study	38

3.6	Analysis of missing data based on Deterministic Method	40
3.7	Analysis of missing data based on Stochastic Modelling	44
3.8	Analysis of missing data based on State Space Method	47
Chapter 4:	Time series analysis and forecasting for stochastic time series data using Box-Jenkins Methodology	
4.1	Preliminary Analysis	49
4.2	Forecasting using En-bloc Methods	52
4.2.1	Forecasting using Decomposition Method	52
4.2.2	Forecasting using Winters' Seasonal Method	57
4.3	Forecasting using Box-Jenkins ARIMA Model	59
4.3.1	Methodology	59
4.3.2	Box-Cox Transformations to stabilize variance	62
4.3.3	Analysis of the data Series	62
4.3.4	Choosing the best Model	70
Chapter 5:	Time series analysis and forecasting for high-volatile data using ARCH/GARCH methodology	
5.1	Introduction	74
5.2	The ARCH Model	74
5.3	Stationarity, Moments, and Restrictions on Parameters	76
5.4	GARCH Modeling	78



5.4.1	Types of volatilities	78
5.4.2	Symmetric GARCH models (Vanila GARCH)	78
5.4.3	Integrated GARCH models	79
5.4.4	Asymmetry Model	80
5.4.5	GJR-GARCH	81
5.4.6	ARCH-IN-MEAN model $\{r\}$	81
5.4.7	NARCH, SQARCH, PARCH, QARCH, STARCH, and APARCH...	82
5.5	Testing for volatility clustering	82
5.5.1	Choosing a volatility model	84
5.5.2	Comparing nested models via tests on parameters	84
5.6	High-Volatile Data set used for the Study	85
5.6.1	Modeling and forecasting values for the Money data series	87
5.6.2	Stationary test on Eviews	88
5.6.3	Transforming non-stationary time series	90
5.6.4	Modelling asset return volatility	93

5.6.5	Predictability of asset return volatilities	94
Chapter 6: Conclusions and Recommendations		
6.1	Introduction	104
6.2	Effectiveness of various approaches and methods	104
6.3	Further comparison between various approaches	106
6.4	Further research direction	108
6.5	Conclusion	108
Reference List		109
Appendices - A	Yield of kurakkan from year 1950-2007 in MT	112
Appendices – B	The SAS output of the missing values for the two seasons YALA and MAHA in the years 1965,1969	113
Appendices C-	The presentation of the data series obtained by filling missing values using state space method	115
Appendices D-	The presentation of the money data series, Source: International Financial Statistics, Central Bank of Sri Lanka	116

## LIST OF FIGURES

	Page
Figure 2.1 Behaviour of an additive model	14
Figure 2.2 Behaviour of an multiplicative model	15
Figure 3.1 Plot of kurakkan series with missing values	40
Figure 3.2 The SAS output of the estimated missing values in the kurakkan Series	42
Figure 3.3 The time series plot of the interpolated values using cubic spline method for the kurakkan series	43
Figure 4.1 Plot of the estimated missing values using cubic spline Method	49
Figure 4.2 Plot of the estimated missing values using state space Method	50
Figure 4.3 Time series plot of the state space data series with the estimated missing values	51
Figure 4.4 Time series decomposition plot of the kurakkan series with estimated missing values	52
Figure 4.5 Seasonal analysis of the kurakkan series with estimated missing values	53
Figure 4.6 Normal probability plot for the kurakkan series	54
Figure 4.7 Time series decomposition for the transformed series	54



Figure 4.8	Normal probability plot for the transformed kurakkan series	56
Figure 4.9	Component analysis for the transformed kurakkan series	56
Figure 4.10	Seasonal analysis for the transformed kurakkan series	57
Figure 4.11	Winters' method plot of the kurakkan series with estimated missing values	58
Figure 4.12	Line graph of the kurakkan series	60
Figure 4.13	ACF and PACF of the kurakkan series with estimated missing values	61
Figure 4.14	The Box-Cox plot for the kurakkan series with estimated missing Values	62
Figure 4.15	The line graph of the transformed kurakkan series	62
Figure 4.16	ACF and PACF of the transformed kurakkan	63
Figure 4.17	Line graph of the 1 <sup>st</sup> differenced kurakkan series	63
Figure 4.18	ACF and PACF of the 1 <sup>st</sup> differenced kurakkan series	64
Figure 4.19	Line graph of the seasonally differenced transformed kurakkan series	65
Figure 4.20	ACF and PACF of the seasonally differenced transformed kurakkan series	65

Figure 4.21	Augmented Dicky-Fuller unit root test of the seasonally differenced transformed kurakkan series	66
Figure 4.22	E-views output of model 1 with 2 AR parameters and 3 MA parameters for the kurakkan series	67
Figure 4.23	ACF and PACF of the residuals of model 1 with 2 AR parameters and 3 MA parameters for the kurakkan series	68
Figure 4.24	E-views output of model 2 with 2 AR parameters and 2 MA parameters for the kurakkan series	69
Figure 4.25	ACF and PACF of the residuals of model 2 with 2 AR parameters and 2 MA parameters for the kurakkan series	70
Figure 5.1	Money Data Series	85
Figure 5.2	The ACF of the money data series	86
Figure 5.3	The PACF of the money data series	87
Figure 5.4	E-views output of the Dickey-Fuller unit root test for the first regression equation	88
Figure 5.5	E-views output of the Dickey-Fuller unit root test for the second regression equation	89
Figure 5.6	The first difference of the money data series	90
Figure 5.7	Line graph of the 1 <sup>st</sup> differenced money series	91
Figure 5.8	E-views output of the 1 <sup>st</sup> differenced money series	91

Figure 5.9	Line graph of the logarithmic transformation of the money series	92
Figure 5.10	The correlogram of the logarithmic transformation of the money Series	92
Figure 5.11	Line graph of the calculated asset return for the money series	93
Figure 5.12	Histogram and descriptive statistics of the asset return for the money series	94
Figure 5.13	E-views output of the asset return for ARMA(1,1) model	95
Figure 5.14	Correlogram of the asset return for ARMA(1,1) model	95
Figure 5.15	Correlogram of the squared residuals for the asset return series	96
Figure 5.16	The output of the ARCH(1) model	97
Figure 5.17	The histogram of the ARCH(1)	98
Figure 5.18	The output of the GARCH(0,1) model	99
Figure 5.19	The histogram of the GARCH(0,1)	100
Figure 5.20	The ARCH-LM test for the GARCH(0,1) model	101
Figure 5.21	The annualized standard deviation for the data series	102
Figure 5.22	The out-of sample forecasts for the money data series	103

## LIST OF TABLES

	Page
Table 2.1      Appropriate values for Box-Cox transformations	20
Table 3.1      Interpolated values for the missing data of the kurakkan series using cubic spline method	36
Table 3.2      The interpolated values using state space modeling	39
Table 4.1      Calculation of MAPE for the decomposition model	49
Table 4.2      Calculation of MAPE for the Winters' seasonal model	50
Table 4.3      The SBC, Adjusted $R^2$ and SE of regression, for all the parsimonious models applied to the transformed kurakkan data series	63
Table 4.4      Estimated forecast values from the fitted model for the transformed kurakkan series	64
Table 4.5      Converted estimated forecast values for the transformed kurakkan series	64
Table 4.6      Calculation of MAPE for the ARIMA model	65
Table 6.1      The MAPE values for the decomposition, Winter's seasonal and ARIMA methods	74

## LIST OF ABBREVIATIONS

Abbreviation	Description
ACF	Auto Correlation Function
ADF	Augmented Dickey Fuller test
AIC	Akaike's Information Criterion
AR	Auto Regressive
ARCH	Autoregressive Conditional Heteroscedastic
ARMA	Auto Regressive Integrated Moving Average
ARIMA	Auto Regressive Integrated Moving Average
DW	Durbin Watson statistic
GARCH	Generalized Autoregressive Conditional Heteroscedastic
IID	Independent and Identically Distributed
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
PAC	Partial Auto Correlation
PACF	Partial Auto Correlation Function
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SBC	Schwarz Bayesian Criterion
WN	White Noise